

CONCENTRATION FOR POISSON U-STATISTICS: SUBGRAPH COUNTS IN RANDOM GEOMETRIC GRAPHS

SASCHA BACHMANN AND MATTHIAS REITZNER

ABSTRACT. Concentration inequalities for subgraph counts in random geometric graphs built over Poisson point processes are proved. The estimates give upper bounds for the probabilities $\mathbb{P}(N \geq M + r)$ and $\mathbb{P}(N \leq M - r)$ where M is either a median or the expectation of a subgraph count N . The bounds for the lower tail have a fast Gaussian decay and the bounds for the upper tail satisfy an optimality condition. A special feature of the presented inequalities is that the underlying Poisson process does not need to have finite intensity measure.

The tail estimates for subgraph counts follow from concentration inequalities for more general local Poisson U-statistics. These bounds are proved using recent general concentration results for Poisson U-statistics and techniques based on the convex distance for Poisson point processes.

1. INTRODUCTION

Random geometric graphs have been a very active area of research for some decades. The most basic model of these graphs is obtained by choosing a random set of vertices in \mathbb{R}^d and connecting any two vertices by an edge whenever their distance does not exceed some fixed parameter $\rho > 0$. In the seminal work [8] by Gilbert, these graphs were suggested as a model for random communication networks. In consequence of the increasing relevance of real-world networks like social networks or wireless networks, variations of Gilbert's model have gained considerable attention in recent years, see e.g. [13, 14, 4, 17, 5]. For a detailed historical overview on the topic, we refer the reader to Penrose's fundamental monograph [18] on random geometric graphs.

For a given random geometric graph \mathfrak{G} and a fixed connected graph H on k vertices, the corresponding *subgraph count* N^H is the random variable that counts the number of occurrences of H as a subgraph of \mathfrak{G} . Note that only non-induced subgraphs are considered throughout the present work. The resulting class of random variables has been studied by many authors, see [18, Chapter 3] for a historical overview on related results.

The purpose of the present paper is to establish *concentration inequalities* for N^H , i.e. upper bounds on the probabilities $\mathbb{P}(N^H \geq M + r)$ and $\mathbb{P}(N^H \leq M - r)$, when the

2010 *Mathematics Subject Classification.* Primary 60D05; Secondary 05C80, 60C05.

Key words and phrases. Random Graphs, Subgraph Counts, Concentration Inequalities, Stochastic Geometry, Poisson Point Process, Convex Distance.

Both authors are partially supported by the German Research Foundation DFG-GRK 1916.

vertices of \mathfrak{G} are given by the points of a Poisson point process. Our concentration results for subgraph counts, gathered in the following theorem, provide estimates for the cases where M is either the expectation or a median of N^H .

Theorem 1.1. *Let η be a Poisson point process in \mathbb{R}^d with non-atomic intensity measure μ . Let H be a connected graph on $k \geq 2$ vertices. Consider the corresponding subgraph count $N^H = N$ in a random geometric graph associated to η . Assume that μ is such that almost surely $N < \infty$. Then all moments of N exist and for all $r \geq 0$,*

$$\begin{aligned} \mathbb{P}(N \geq \mathbb{E}N + r) &\leq \exp \left(-\frac{((\mathbb{E}N + r)^{1/(2k)} - (\mathbb{E}N)^{1/(2k)})^2}{2k^2 c_d} \right), \\ \mathbb{P}(N \leq \mathbb{E}N - r) &\leq \exp \left(-\frac{r^2}{2k \mathbb{V}N} \right), \\ \mathbb{P}(N > \mathbb{M}N + r) &\leq 2 \exp \left(-\frac{r^2}{4k^2 c_d (r + \mathbb{M}N)^{2-1/k}} \right), \\ \mathbb{P}(N < \mathbb{M}N - r) &\leq 2 \exp \left(-\frac{r^2}{4k^2 c_d (\mathbb{M}N)^{2-1/k}} \right), \end{aligned}$$

where $c_d > 0$ is a constant that depends on H and d only, $\mathbb{E}N$ and $\mathbb{V}N$ denote the expectation and the variance of N , and $\mathbb{M}N$ is the smallest median of N .

One might think that the study of subgraph counts needs to be restricted to finite graphs to ensure that all occurring variables are finite. We stress that this is *not* the case. There are Poisson processes in \mathbb{R}^d , such that the associated random geometric graph has a.s. infinitely many vertices but still a.s. a finite number of edges. Similarly, one can also have a.s. infinitely many edges, but still a finite number of triangles. This phenomenon seems to be quite unexplored so far. In the context of concentration properties, a natural question is whether concentration inequalities also hold in these situations. We emphasize that Theorem 1.1 only requires $N < \infty$ almost surely and hence covers such cases, as opposed to previous results from [21, 15] where only finite intensity measures are considered.

The asymptotic exponents in the upper tail bounds for both the expectation and the median are equal to $1/k$. This is actually best possible as will be pointed out later. Note also that the asymptotic exponent of the estimates in previous results from [21, 15] are $1/k$ for both the upper and the lower tail which is compatible with our upper tail bounds but worse than our lower tail inequalities.

The essential tools for proving our concentration inequalities are new results on Poisson U-statistics. Given a (simple) Poisson point process η on \mathbb{R}^d , denote by η_{\neq}^k the set of all k -tuples of distinct elements of η . A *Poisson U-statistic* of order k is a random variable that can be written as

$$\sum_{\mathbf{x} \in \eta_{\neq}^k} f(\mathbf{x}),$$

where $f : (\mathbb{R}^d)^k \rightarrow \mathbb{R}$ is a symmetric measurable map. The methods that we use in the present work to prove concentration of Poisson U-statistics around the mean are based on tools recently developed by Bachmann and Peccati in [1]. In this work, concentration inequalities for the edge count in random geometric graphs were established. The approach to achieve these estimates will be generalized in the present article to arbitrary subgraph counts.

In addition to this, we will present a refinement of the method suggested in [21, 15] by Lachièze-Rey, Reitzner, Schulte and Thäle which gives concentration of Poisson U-statistics around the median. We improve this method further with the particular advantage, apart from giving clean and easy to use tail bounds, that the underlying Poisson process may have infinite intensity measure.

As an application of the concentration inequalities in Theorem 1.1 we are going to establish strong laws of large numbers for subgraph counts. In [18, Chapter 3] one can find strong laws for subgraph counts associated to i.i.d. points in \mathbb{R}^d . To the best of our knowledge, there are no comparable strong laws so far for the Poisson process case and a particular feature of our results is that they even apply to certain Poisson processes with non-finite intensity measure.

In recent years, the study of random geometric simplicial complexes built on random point sets in \mathbb{R}^d has attracted a considerable attention, see e.g. [10, 6, 11, 24] and in particular the survey [2] by Bobrowski and Kahle on this topic. Motivations to study these complexes arise particularly from topological data analysis (for a survey on this see [3]), but also from other applications of geometric complexes like sensor networks (see e.g. [23]). One of the models for random simplicial complexes is the so-called Vietoris-Rips complex. We stress that the number of k -dimensional simplices in this complex is exactly given by the number of complete subgraphs on k -vertices in the corresponding random geometric graph. So the results in the present paper are directly related to the study of random Vietoris-Rips complexes and may be useful in future research on these models.

Related to our investigations are the findings by Eichelsbacher, Raič and Schreiber in [7] where deviation inequalities for stabilizing functionals of finite intensity measure Poisson processes were derived. These results are in principle also applicable to U-statistics (and hence to subgraph counts), as it was pointed out in [21], although the appearing asymptotic exponents tend to be non-optimal. Moreover, since the constants in the tail bounds from [7] depend on the intensity and are not given explicitly, it remains as an interesting open question whether these estimates can be extended to settings with non-finite intensity measures.

The paper is organized as follows. In Section 2 we will introduce the general assumptions and definitions that form the framework for the entire work. As anticipated, we will prove the announced concentration inequalities not only for subgraph counts, but for local Poisson U-statistics that satisfy certain additional assumptions. The tail estimates for subgraph counts will then follow directly from the more general results. In Section 3 we will present these general concentration inequalities. The

announced applications to subgraph counts in random geometric graphs are presented in Section 4. Here, after pointing out how the general results apply to the setting of subgraph counts, we will analyse the asymptotic behaviour of expectation, median and variance which is needed for the subsequent presentation of the strong laws of large numbers. The proofs of all statements are gathered in Section 5.

2. FRAMEWORK

For the remainder of the article we denote by \mathbf{N} the space of locally finite point configurations in \mathbb{R}^d . Elements in \mathbf{N} can be regarded as locally finite subsets of \mathbb{R}^d , but also as locally finite simple point measures on \mathbb{R}^d . In this spirit, for $\xi \in \mathbf{N}$, we will use set related notations like for example $\xi \cap A$ as well as measure related notations like $\xi(A)$. Moreover, we denote by \mathcal{N} the σ -algebra over \mathbf{N} that is generated by the maps

$$\mathbf{N} \rightarrow \mathbb{N} \cup \{\infty\}, \xi \mapsto \xi(A),$$

where A ranges over all Borel subsets of \mathbb{R}^d . Throughout, we will consider a (non-trivial) Poisson point process η on \mathbb{R}^d . Then η is a random element in \mathbf{N} and its intensity measure μ is the measure on \mathbb{R}^d defined by $\mu(A) = \mathbb{E}\eta(A)$ for any Borel set $A \subseteq \mathbb{R}^d$. It will be assumed that μ is locally finite and does not have atoms. A (*Poisson*) *U-statistic of order k* is a functional $F : \mathbf{N} \rightarrow \mathbb{R} \cup \{\pm\infty\}$ that can be written as

$$F(\xi) = \sum_{\mathbf{x} \in \xi_{\neq}^k} f(\mathbf{x}),$$

where $f : (\mathbb{R}^d)^k \rightarrow \mathbb{R}$ is a symmetric measurable map called the *kernel* of F , and for any $\xi \in \mathbf{N}$,

$$\xi_{\neq}^k = \{(x_1, \dots, x_k) \in \xi^k : x_i \neq x_j \text{ whenever } i \neq j\}.$$

The map F and the corresponding random variable $F(\eta)$ will be identified in what follows if there is no risk of ambiguity. Note that in this work we will only consider U-statistics F with non-negative kernels. Also, it will be assumed throughout that η guarantees almost surely $F(\eta) < \infty$ if not stated otherwise.

Motivated by the application to random geometric graphs, we will define further properties that a U-statistic F with kernel $f \geq 0$ might satisfy. We will refer to these properties in the remainder of the article.

(K1) There is a constant $\rho_F > 0$ such that

$$f(x_1, \dots, x_k) > 0 \text{ whenever } \text{diam}(x_1, \dots, x_k) \leq \rho_F.$$

(K2) There is a constant $\Theta_F \geq 1$ such that

$$f(x_1, \dots, x_k) = 0 \text{ whenever } \text{diam}(x_1, \dots, x_k) > \Theta_F \rho_F.$$

(K3) There are constants $M_F \geq m_F > 0$ such that

$$M_F \geq f(x_1, \dots, x_k) \geq m_F \text{ whenever } f(x_1, \dots, x_k) > 0.$$

U-Statistics that satisfy property (K2) are also referred to as *local* U-Statistics. Property (K3) is particularly satisfied for U-Statistics that count occurrences of certain subconfigurations.

3. GENERAL RESULTS

The concentration inequalities for U-statistics we are about to present are based on methods developed in [21, 15] and [1]. We begin by citing the results that we use from these articles. To do so, we first need to introduce a further notion. The *local version* of a U-statistic F is defined for any $\xi \in \mathbf{N}$ and $x \in \xi$ by

$$F(x, \xi) = \sum_{\mathbf{y} \in (\xi \setminus x)^{k-1}} f(x, \mathbf{y}),$$

where $\xi \setminus x$ is shorthand for $\xi \setminus \{x\}$. Note, that

$$F(\xi) = \sum_{x \in \xi} F(x, \xi).$$

The upcoming result uses a notion introduced in [1]: A U-statistic F is called *well-behaved* if there is a measurable set $B \subseteq \mathbf{N}$ with $\mathbb{P}(\eta \in B) = 1$, such that

- (i) $F(\xi) < \infty$ for all $\xi \in B$,
- (ii) $\xi \cup \{x\} \in B$ whenever $\xi \in B$ and $x \in \mathbb{R}^d$,
- (iii) $\xi \setminus \{x\} \in B$ whenever $\xi \in B$ and $x \in \xi$.

Roughly speaking, the motivation for this notion is that for a well-behaved U-statistic F one has almost surely $F(\eta) < \infty$ as well as $F(\eta \cup \{x\}) < \infty$ for $x \in \mathbb{R}^d$ and $F(\eta \setminus \{x\}) < \infty$ for $x \in \eta$. This ensures that the local versions $F(x, \eta)$ for $x \in \eta$ and also $F(x, \eta \cup \{x\})$ for $x \in \mathbb{R}^d$ are almost surely finite, thus preventing technical problems arising from non-finiteness of these quantities. The following theorem combines [1, Corollary 5.2] and [1, Corollary 5.3] and the fact that by [12, Lemma 3.4] the exponential decay of the upper tail implies existence of all moments of F .

Theorem 3.1. *Let F be a well-behaved U-statistic of order k with kernel $f \geq 0$. Assume that for some $\alpha \in [0, 2)$ and $c > 0$ we have almost surely*

$$(3.1) \quad \sum_{x \in \eta} F(x, \eta)^2 \leq cF^\alpha.$$

Then all moments of F exist and for all $r \geq 0$,

$$\mathbb{P}(F \geq \mathbb{E}F + r) \leq \exp \left(-\frac{((\mathbb{E}F + r)^{1-\alpha/2} - (\mathbb{E}F)^{1-\alpha/2})^2}{2k^2c} \right)$$

and

$$\mathbb{P}(F \leq \mathbb{E}F - r) \leq \exp \left(-\frac{r^2}{2k\mathbb{V}F} \right).$$

Note that condition (K2), as defined in Section 2, clearly ensures that all U-statistics considered in the present paper are well-behaved in the sense described above.

In addition to the latter result, a variation of the approach presented in [21, 15] gives that for finite intensity measure Poisson processes, the condition (3.1) also implies a concentration inequality for the median instead of the expectation of the considered U-statistic. Moreover, it is possible to extend these tail estimates to U-statistics built over non-finite intensity measure processes, resulting in the forthcoming theorem. To state this result, we first need to introduce a further notation. For any real random variable Z , we denote by $\mathbf{M}Z$ the smallest median of Z , i.e.

$$(3.2) \quad \mathbf{M}Z = \inf\{x \in \mathbb{R} : \mathbb{P}(Z \leq x) \geq 1/2\}.$$

Note that $\mathbf{M}Z$ is exactly the value which the quantile-function of Z takes at $1/2$. We are well prepared to state the announced result.

Theorem 3.2. *Let F be a U-statistic of order k with kernel $f \geq 0$. Assume that F is almost surely finite and satisfies (3.1) for some $\alpha \in [0, 2)$ and $c > 0$. Then for all $r \geq 0$,*

$$\begin{aligned} \mathbb{P}(F > \mathbf{M}F + r) &\leq 2 \exp\left(-\frac{r^2}{4k^2c(r + \mathbf{M}F)^\alpha}\right), \\ \mathbb{P}(F < \mathbf{M}F - r) &\leq 2 \exp\left(-\frac{r^2}{4k^2c(\mathbf{M}F)^\alpha}\right). \end{aligned}$$

In the light of the above results, a natural approach towards concentration inequalities for U-statistics is to establish condition (3.1). For U-statistics satisfying the conditions (K1) to (K3) we obtain the following.

Theorem 3.3. *Let F be a U-statistic of order k with kernel $f \geq 0$ that satisfies (K1) to (K3). Then for any $\xi \in \mathbf{N}$ we have*

$$(3.3) \quad \sum_{x \in \xi} F(x, \xi)^2 \leq c_d F(\xi)^{\frac{2k-1}{k}},$$

where

$$c_d = \left(2 \left\lceil \Theta_F \sqrt{d} \right\rceil + 1\right)^{2d(k-1)} \cdot M_F^2 \left(\frac{k^k}{m_F k!}\right)^{\frac{2k-1}{k}}.$$

The above statement guarantees that Theorem 3.1 and Theorem 3.2 apply to any almost surely finite U-statistic F that satisfies (K1) to (K3). Thus, the following tail estimates are established.

Corollary 3.4. *Let F be a U-statistic of order k with kernel $f \geq 0$. Assume that F is almost surely finite and satisfies (K1) to (K3). Then all moments of F exist and for all $r \geq 0$,*

$$\begin{aligned} \mathbb{P}(F \geq \mathbb{E}F + r) &\leq \exp\left(-\frac{((\mathbb{E}F + r)^{1/(2k)} - (\mathbb{E}F)^{1/(2k)})^2}{2k^2 c_d}\right), \\ \mathbb{P}(F \leq \mathbb{E}F - r) &\leq \exp\left(-\frac{r^2}{2k \mathbb{V}F}\right), \\ \mathbb{P}(F > \mathbb{M}F + r) &\leq 2 \exp\left(-\frac{r^2}{4k^2 c_d (r + \mathbb{M}F)^{2-1/k}}\right), \\ \mathbb{P}(F < \mathbb{M}F - r) &\leq 2 \exp\left(-\frac{r^2}{4k^2 c_d (\mathbb{M}F)^{2-1/k}}\right), \end{aligned}$$

where c_d is defined as in Theorem 3.3.

We conclude this section with a brief discussion about optimality of the upper tail concentration inequalities that were established above. Consider a U-statistic F such that the assumptions of Corollary 3.4 hold. Assume that

$$(3.4) \quad \mathbb{P}(F \geq M + r) \leq \exp(-I(r)),$$

where M is either the mean or a median of F and I is a function that satisfies

$$(3.5) \quad \liminf_{r \rightarrow \infty} I(r)/r^a > 0$$

for some $a > 0$. The upper tail estimates from Corollary 3.4 yield such functions I for the exponent $a = 1/k$. The next result states that this is optimal.

Proposition 3.5. *Let F be a U-statistic of order k with positive kernel $f \geq 0$ such that F satisfies (K1) to (K3). Let I and $a > 0$ be such that (3.4) and (3.5) are verified. Then $a \leq 1/k$.*

4. SUBGRAPH COUNTS IN RANDOM GEOMETRIC GRAPHS

In the following, we are going to investigate a model for random geometric graphs which was particularly investigated in [13] and [14]. Let $S \subseteq \mathbb{R}^d$ be such that $S = -S$. To any countable subset $\xi \subset \mathbb{R}^d$ we assign the *geometric graph* $\mathfrak{G}_S(\xi)$ with vertex set ξ and an edge between two distinct vertices x and y whenever $x - y \in S$. For $\mathbf{x} = (x_1, \dots, x_k) \in (\mathbb{R}^d)^k$ we will occasionally write $\mathfrak{G}_S(\mathbf{x})$ instead of $\mathfrak{G}_S(\{x_1, \dots, x_k\})$. Now, assuming that the vertices are chosen at random according to some Poisson point process η , we obtain the *random geometric graph* $\mathfrak{G}_S(\eta)$. Denote the closed ball centered at $x \in \mathbb{R}^d$ with radius $\rho \in \mathbb{R}_+$ by $B(x, \rho)$. Throughout, we will assume that $B(0, \rho) \subseteq S \subseteq B(0, \theta\rho)$ for some $\rho > 0$ and $\theta \geq 1$. Note that if we take $\theta = 1$, then $S = B(0, \rho)$ and we end up with the classical model of random geometric graphs for the euclidean norm, often referred to as *random disk graph*. Also, the classical geometric graphs based on any other norm in \mathbb{R}^d are covered by the model introduced above. These classical models are extensively described in [18] for the case when the underlying point process η has finite intensity measure.

Before proceeding with the discussion, we present a picture that illustrates how the graphs that will be considered in the following might look like (in a window around the origin).

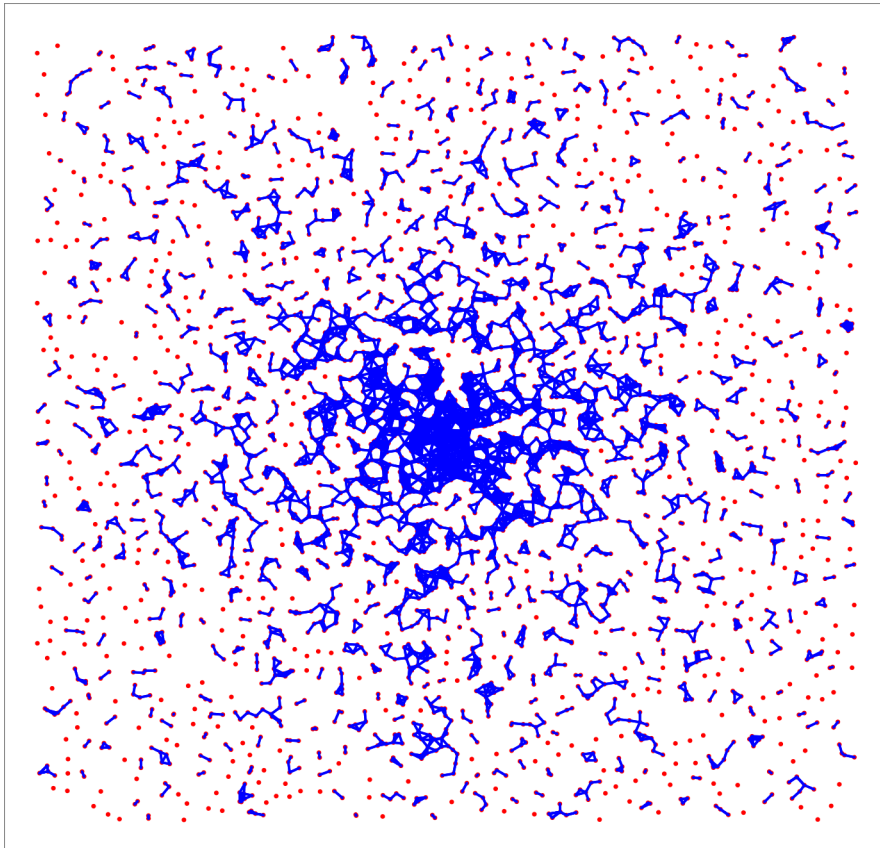


FIGURE 1. Random unit disk graph, intensity measure $\mu = 18(\|x\| + 1)^{-1}dx$

A class of random variables that is frequently studied in the literature on random geometric graphs are the subgraph counts. For any connected graph H on k vertices the corresponding *subgraph count* is the U-statistic

$$N^H(\eta) = \sum_{\mathbf{x} \in \eta_{\neq}^k} f_H(\mathbf{x}),$$

where the kernel f_H is given by

$$f_H(\mathbf{x}) = \frac{|\{\text{subgraphs } H' \text{ of } \mathfrak{G}_S(\mathbf{x}) : H' \cong H\}|}{k!}.$$

At this, a subgraph H' of $\mathfrak{G}_S(\mathbf{x})$ is a graph on a vertex set which is a subset of $\{x_1, \dots, x_k\}$ such that any edge of H' is also an edge of $\mathfrak{G}_S(\mathbf{x})$. Note that we write $H' \cong H$ if H' and H are isomorphic graphs. So, the random variable $N^H(\eta)$ counts the occurrences of H as a subgraph of $\mathfrak{G}_S(\eta)$. We will write $N = N^H$ if H is clear from the context. Note that we do not consider induced subgraphs here. Also

note that since we are particularly interested in graphs built on non-finite intensity measure Poisson processes, it is possible that N takes the value ∞ .

4.1. Concentration Inequalities for Subgraph Counts. From the general concentration result Corollary 3.4 we obtain the deviation inequalities for subgraph counts stated in Theorem 1.1. To see this, let H be a connected graph on k vertices and consider the corresponding subgraph count N^H . Denote by a_H the number of subgraphs isomorphic to H in a complete graph on k vertices. Moreover, let $\text{diam}(H)$ be the diameter of H , i.e. the length of the longest shortest path in H . Then $N = N^H$ satisfies the conditions (K1) to (K3) with

$$\rho_N = \rho, \quad \Theta_N = \text{diam}(H)\theta, \quad M_N = \frac{a_H}{k!}, \quad m_N = \frac{1}{k!}.$$

Hence, Theorem 1.1 follows from Corollary 3.4 where

$$c_d = \left(2 \left\lceil \text{diam}(H)\theta\sqrt{d} \right\rceil + 1\right)^{2d(k-1)} \cdot \left(\frac{a_H}{k!}\right)^2 k^{2k-1}.$$

4.2. Asymptotic Behaviour of Subgraph Counts. The concentration inequalities presented in our main Theorem 1.1 depend on expectation, median or variance of the considered subgraph count. It is therefore of interest, how these quantities behave asymptotically in settings where the parameters of the model are being varied. In what comes, we will study these asymptotic behaviours for sequences of subgraph counts with the particular goal of establishing strong laws of large numbers.

For the remainder of this section, we consider a sequence of random geometric graphs $(\mathfrak{G}_{\rho_t S}(\eta_t))_{t \in \mathbb{N}}$. At this, $(\eta_t)_{t \in \mathbb{N}}$ is a sequence of Poisson point processes in \mathbb{R}^d where each η_t has intensity measure $\mu_t = t\mu$ and μ is given by a Lebesgue density m . Moreover, $(\rho_t)_{t \in \mathbb{N}}$ is a sequence of positive real numbers such that $\rho_t \rightarrow 0$ as $t \rightarrow \infty$. In correspondence with the conventions at the beginning of Section 4, the set S is assumed to satisfy $B(0, 1) \subseteq S \subseteq B(0, \theta)$ for some $\theta \geq 1$. Any connected graph H on k vertices now yields a corresponding sequence $N_t^H = N_t$ of subgraph counts in the random graphs $\mathfrak{G}_{\rho_t S}(\eta_t)$.

We are particularly interested in settings where the underlying Poisson point process has non-finite intensity measure. In this situation it is not automatically guaranteed that the considered subgraph counts are almost surely finite. Note (again) that by Theorem 1.1, almost sure finiteness of a subgraph count is equivalent to existence of all its moments. The upcoming result gives a sufficient condition for almost sure finiteness of the considered random variables.

Proposition 4.1. *Assume that the Lebesgue density m is bounded. Assume in addition that there exists a constant $c \geq 1$ such that*

$$(4.6) \quad m(x) \leq cm(y) \quad \text{whenever} \quad \|x - y\| \leq \text{diam}(H)\theta \sup_{t \in \mathbb{N}} \rho_t.$$

Then, for any $t \in \mathbb{N}$, the random variable N_t is almost surely finite if and only if

$$(4.7) \quad \int_{\mathbb{R}^d} m(x)^k dx < \infty.$$

Remark. The reason for considering densities that satisfy condition (4.6) is that this allows us to use a dominated convergence argument in the proof of the upcoming Theorem 4.2. A class of densities that satisfy condition (4.6) is given by

$$m(x) = A(\|x\| + 1)^{-\gamma},$$

where $A, \gamma > 0$. Indeed, for any $x, y \in \mathbb{R}^d$,

$$\frac{m(x)}{m(y)} = \left(\frac{\|y\| + 1}{\|x\| + 1} \right)^\gamma \leq \left(\frac{\|x\| + \|x - y\| + 1}{\|x\| + 1} \right)^\gamma \leq (1 + \|x - y\|)^\gamma.$$

An example of how the resulting random graphs for these densities might look like is illustrated in Figure 1 above. Observe also that condition (4.7) for a density m given by $m(x) = A(\|x\| + 1)^{-\gamma}$ is equivalent to $k > d/\gamma$. In particular, the graph in Figure 1 has almost surely infinitely many edges, but any subgraph count of order at least 3 (for example the number of triangles) is almost surely finite.

Our goal is to establish strong laws for suitably rescaled versions of the subgraph counts N_t . To do so, it is crucial to know the asymptotic behaviour of the expectation $\mathbb{E}N_t$ and the variance $\mathbb{V}N_t$ as $t \rightarrow \infty$. In the case when the intensity measure of the underlying Poisson point process is finite, these asymptotics are well known (see e.g. [18]). Since our focus is on random geometric graphs built on non-finite intensity measure processes, the first step towards establishing strong laws is to study expectation and variance asymptotics. We also have concentration inequalities with respect to the median, so the asymptotic behaviour of the median may as well be of interest in applications. The upcoming result addresses these issues. Note that we will write $a_n \sim b_n$ if sequences $(a_n)_{n \in \mathbb{N}}$ and $(b_n)_{n \in \mathbb{N}}$ are *asymptotically equivalent*, meaning that $\lim_{n \rightarrow \infty} a_n/b_n = 1$.

Theorem 4.2. *Assume that the intensity measure μ is given by a continuous and bounded Lebesgue density $m : \mathbb{R}^d \rightarrow \mathbb{R}_+$ that satisfies (4.6) and (4.7). Then the following holds:*

- (i) *There exists a constant $a > 0$ such that*

$$\mathbb{E}N_t \sim at^k \rho_t^{d(k-1)}.$$

- (ii) *If $\lim_{t \rightarrow \infty} \mathbb{E}N_t = \infty$, then*

$$\mathbb{M}N_t \sim \mathbb{E}N_t.$$

- (iii) *There exist constants $A^{(n)} > 0$ for $1 \leq n \leq k$ such that*

$$\mathbb{V}N_t \sim t^k \rho_t^{d(k-1)} \sum_{n=1}^k (t \rho_t^d)^{k-n} A^{(n)}.$$

The concentration inequalities together with asymptotic results for expectation and variance can be used to obtain the upcoming strong laws for subgraph counts. Note that for the next result to hold, we only need that expectation and variance behave asymptotically as stated in Theorem 4.2. According to [18, Proposition 3.1, Proposition 3.7] this asymptotic behaviour is also in order whenever the intensity

measure of η is finite and has a bounded Lebesgue density, so the following theorem also applies in these situations. The strong law presented below complements the results [18, Theorem 3.17, Theorem 3.18] that deal with random geometric graphs built over i.i.d. points.

Theorem 4.3. *Assume that the statements (i) and (iii) of Theorem 4.2 hold. Assume in addition that for some $\gamma > 0$,*

$$(4.8) \quad \liminf_{t \rightarrow \infty} t^{k-\gamma} \rho_t^{d(k-1)} > 0.$$

Let $a > 0$ denote the limit of the sequence $\mathbb{E}N_t/(t^k \rho_t^{d(k-1)})$. Then

$$\frac{N_t}{t^k \rho_t^{d(k-1)}} \xrightarrow{a.s.} a \quad \text{as } t \rightarrow \infty.$$

5. PROOFS

Proof of Theorem 3.3. To get prepared for the proof of Theorem 3.3, which is crucial to establish the desired deviation and concentration inequalities, we first prove the following two lemmas.

Lemma 5.1. *Let $x_1, \dots, x_k, y_1, \dots, y_k \in \mathbb{R}_{\geq 0}$. Then*

$$\prod_{i=1}^k x_i + \prod_{i=1}^k y_i \leq \prod_{i=1}^k \max(x_i, y_i) + \prod_{i=1}^k \min(x_i, y_i).$$

Proof. Let I and J be the sets of indices i satisfying $x_i < y_i$ and $x_i \geq y_i$, respectively. Then we have

$$\prod_{i \in I} x_i < \prod_{i \in I} y_i \quad \text{and} \quad \prod_{i \in J} x_i \geq \prod_{i \in J} y_i.$$

It follows that

$$\begin{aligned} \prod_{i=1}^k x_i + \prod_{i=1}^k y_i &= \prod_{i \in I} x_i \prod_{i \in J} x_i + \prod_{i \in I} y_i \prod_{i \in J} y_i \\ &= \prod_{i \in I} x_i \prod_{i \in J} x_i + \left(\prod_{i \in I} y_i - \prod_{i \in I} x_i \right) \prod_{i \in J} y_i + \prod_{i \in I} x_i \prod_{i \in J} y_i \\ &\leq \prod_{i \in I} x_i \prod_{i \in J} x_i + \left(\prod_{i \in I} y_i - \prod_{i \in I} x_i \right) \prod_{i \in J} x_i + \prod_{i \in I} x_i \prod_{i \in J} y_i \\ &= \prod_{i \in I} y_i \prod_{i \in J} x_i + \prod_{i \in I} x_i \prod_{i \in J} y_i = \prod_{i=1}^k \max(x_i, y_i) + \prod_{i=1}^k \min(x_i, y_i). \end{aligned}$$

□

Using the lemma above, we can prove the following.

Lemma 5.2. *Let $n_1, \dots, n_N \in \mathbb{R}_{\geq 0}$, $k \geq 2$ and let π_1, \dots, π_k be permutations of $\{1, \dots, N\}$. Then*

$$\sum_{i=1}^N \prod_{j=1}^k n_{\pi_j(i)} \leq \sum_{i=1}^N n_i^k.$$

Proof. Without loss of generality we can assume

$$(5.9) \quad n_1 \leq n_2 \leq \dots \leq n_N.$$

The proof is by induction on N . For $N = 1$ there is nothing to prove, for $N = 2$ the result follows from Lemma 5.1. So let $N > 2$ and assume the result holds for $N - 1$. Let

$$t = |\{i : \pi_j(i) = N \text{ for some } j\}|.$$

We prove by induction on t that the following holds: There exist permutations $\tilde{\pi}_1, \dots, \tilde{\pi}_k$ of $\{1, \dots, N - 1\}$ such that

$$\sum_{i=1}^N \prod_{j=1}^k n_{\pi_j(i)} \leq \sum_{i=1}^{N-1} \prod_{j=1}^k n_{\tilde{\pi}_j(i)} + n_N^k.$$

The result then follows by applying the induction hypotheses (with respect to N) to the right hand side of this inequality.

For $t = 1$ the permutations $\tilde{\pi}_i$ obviously exist since in this case there is an index l such that $\pi_j(l) = N$ for all j . Consider the case $t > 1$. Then there exist indices $l_1 \neq l_2$ such that $\pi_{j_1}(l_1) = \pi_{j_2}(l_2) = N$ for some j_1, j_2 . We define permutations $\bar{\pi}_1, \dots, \bar{\pi}_k$ of $\{1, \dots, N\}$ by $\bar{\pi}_j = \pi_j$ if $\pi_j(l_1) > \pi_j(l_2)$ and by

$$\bar{\pi}_j(i) = \begin{cases} \pi_j(i) & \text{if } i \neq l_1, l_2 \\ \pi_j(l_2) & \text{if } i = l_1 \\ \pi_j(l_1) & \text{if } i = l_2 \end{cases}$$

if $\pi_j(l_1) < \pi_j(l_2)$. Then, using Lemma 5.1 together with (5.9) we obtain

$$\begin{aligned} \sum_{i=1}^N \prod_{j=1}^k n_{\pi_j(i)} &= \sum_{i=1, i \neq l_1, l_2}^N \prod_{j=1}^k n_{\pi_j(i)} + \prod_{j=1}^k n_{\pi_j(l_1)} + \prod_{j=1}^k n_{\pi_j(l_2)} \\ &\leq \sum_{i=1, i \neq l_1, l_2}^N \prod_{j=1}^k n_{\pi_j(i)} + \prod_{j=1}^k \max(n_{\pi_j(l_1)}, n_{\pi_j(l_2)}) + \prod_{j=1}^k \min(n_{\pi_j(l_1)}, n_{\pi_j(l_2)}) \\ &= \sum_{i=1}^N \prod_{j=1}^k n_{\bar{\pi}_j(i)}. \end{aligned}$$

Note, that $\bar{\pi}_j(l_2) \neq N$ for all j and hence $|\{i : \bar{\pi}_j(i) = N \text{ for some } j\}| = t - 1$. Applying the induction hypotheses (with respect to t) yields the existence of the desired permutations $\tilde{\pi}_j$. This concludes the proof. \square

Proof of Theorem 3.3. First note that the result holds trivially in the case when $F(\xi) = \infty$. Moreover, if $F(\xi) < \infty$, then property (K3) guarantees that $F(x, \xi) = 0$ for all but finitely many $x \in \xi$. Hence, we can assume without loss of generality that ξ is finite and that $F(x, \xi) > 0$ for all $x \in \xi$.

We consider a tiling of \mathbb{R}^d into cubes of diagonal ρ_F . Since we assumed that ξ is finite, we can choose out of these cubes C_1, \dots, C_N such that $\xi \subseteq \cup C_i$. Then for any distinct $x_1, \dots, x_k \in \xi$ contained in the same cube we have $f(x_1, \dots, x_k) \geq m_F$. Moreover, translating the tiling a bit if necessary, we can assume without loss of generality that each point $x \in \xi$ is contained in the interior of exactly one of the cubes.

Observe that condition (K2) implies the following: Let $i \in \{1, \dots, N\}$. Then there are $(2\lceil\sqrt{d}\Theta_F\rceil + 1)^d =: q$ many cubes C_i^1, \dots, C_i^q such that for any $x \in C_i$ we have $\{y_1, \dots, y_{k-1}\} \subset \cup_{j=1}^q C_i^j$ whenever $f(x, y_1, \dots, y_{k-1}) > 0$. Note also, that the C_i^j can be chosen such that for any fixed j it holds that $\{C_1, \dots, C_N\} = \{C_1^j, \dots, C_N^j\}$. Let $[q] := \{1, \dots, q\}$ and for any $x \in \xi \cap C_i$ and $(j_1, \dots, j_{k-1}) \in [q]^{k-1}$ let

$$C_i(x, j_1, \dots, j_{k-1}) := \left[(C_i^{j_1} \cap (\xi \setminus x)) \times \dots \times (C_i^{j_{k-1}} \cap (\xi \setminus x)) \right]_{\neq}.$$

Then we have

$$\begin{aligned} \sum_{x \in \xi} F(x, \xi)^2 &= \sum_{i=1}^N \sum_{x \in \xi \cap C_i} F(x, \xi)^2 \\ &= \sum_{i=1}^N \sum_{x \in \xi \cap C_i} \left(\sum_{(j_1, \dots, j_{k-1}) \in [q]^{k-1}} \sum_{\mathbf{y} \in C_i(x, j_1, \dots, j_{k-1})} f(x, \mathbf{y}) \right)^2, \end{aligned}$$

where $\mathbf{y} = (y_1, \dots, y_{k-1})$. Now let $n_i = |\xi \cap C_i|$ and $n_i^{(j)} = |\xi \cap C_i^j|$. Then by condition (K3) the last expression in the above display does not exceed

$$\sum_{i=1}^N \sum_{x \in \xi \cap C_i} \left(\sum_{(j_1, \dots, j_{k-1}) \in [q]^{k-1}} M_F \prod_{l=1}^{k-1} n_i^{(j_l)} \right)^2.$$

Moreover, we have

$$\left(\sum_{(j_1, \dots, j_{k-1}) \in [q]^{k-1}} M_F \prod_{l=1}^{k-1} n_i^{(j_l)} \right)^2 \leq q^{k-1} M_F^2 \sum_{(j_1, \dots, j_{k-1}) \in [q]^{k-1}} \prod_{l=1}^{k-1} (n_i^{(j_l)})^2.$$

Thus, it follows from the above considerations that

$$\sum_{x \in \xi} F(x, \xi)^2 \leq q^{k-1} M_F^2 \sum_{i=1}^N \sum_{x \in \xi \cap C_i} \sum_{(j_1, \dots, j_{k-1}) \in [q]^{k-1}} \prod_{l=1}^{k-1} (n_i^{(j_l)})^2.$$

Rearranging the triple sum, we can write the right hand side as

$$\begin{aligned} & q^{k-1} M_F^2 \sum_{(j_1, \dots, j_{k-1}) \in [q]^{k-1}} \sum_{i=1}^N \sum_{x \in \xi \cap C_i} \prod_{l=1}^{k-1} (n_i^{(j_l)})^2 \\ &= q^{k-1} M_F^2 \sum_{(j_1, \dots, j_{k-1}) \in [q]^{k-1}} \sum_{i=1}^N n_i \prod_{l=1}^{k-1} (n_i^{(j_l)})^2. \end{aligned}$$

By Lemma 5.2 this expression does not exceed

$$\begin{aligned} q^{k-1} M_F^2 \sum_{(j_1, \dots, j_{k-1}) \in [q]^{k-1}} \sum_{i=1}^N n_i^{2k-1} &= q^{2(k-1)} M_F^2 \sum_{i=1}^N n_i^{2k-1} \\ &\leq q^{2(k-1)} M_F^2 \left(\sum_{i=1}^N n_i^k \right)^{\frac{2k-1}{k}}. \end{aligned}$$

At this, the latter inequality holds by monotonicity of the p -Norm.

It remains to prove

$$\sum_{i=1}^N n_i^k \leq \frac{k^{k-1}}{m_F(k-1)!} \sum_{x \in \xi} F(x, \xi).$$

We will see that for any $x \in \xi \cap C_i$ we have

$$(5.10) \quad n_i^{k-1} \frac{(k-1)!}{k^{k-1}} \leq \frac{F(x, \xi)}{m_F}.$$

This then gives, as desired, that

$$\begin{aligned} \frac{k^{k-1}}{(k-1)!} \sum_{x \in \xi} F(x, \xi) &= m_F \sum_{i=1}^N \sum_{x \in \xi \cap C_i} \frac{k^{k-1} F(x, \xi)}{m_F(k-1)!} \\ &\geq m_F \sum_{i=1}^N \sum_{x \in \xi \cap C_i} n_i^{k-1} = m_F \sum_{i=1}^N n_i^k. \end{aligned}$$

To prove (5.10), let $x \in \xi \cap C_i$. Consider the set

$$A := \{(y_1, \dots, y_{k-1}) \in (\xi \setminus x)^{k-1} : f(x, y_1, \dots, y_{k-1}) > 0\}.$$

Then, by definition of $F(x, \xi)$ and by condition (K3) we have

$$F(x, \xi) = \sum_{(y_1, \dots, y_{k-1}) \in A} f(x, y_1, \dots, y_{k-1}) \geq \sum_{(y_1, \dots, y_{k-1}) \in A} m_F.$$

Thus

$$\frac{F(x, \xi)}{m_F} \geq |A|.$$

Since by condition (K1) we have $((\xi \setminus x) \cap C_i)_{\neq}^{k-1} \subseteq A$, it follows that

$$\prod_{t=1}^{k-1} (n_i - t) = \prod_{t=0}^{k-2} (n_i - 1 - t) = |((\xi \setminus x) \cap C_i)_{\neq}^{k-1}| \leq \frac{F(x, \xi)}{m_F}.$$

Moreover, it is straightforward to check that

$$\frac{\prod_{t=1}^{k-1} (n_i - t)}{n_i^{k-1}} \geq \frac{(k-1)!}{k^{k-1}}$$

whenever $n_i > k$. Hence, it follows that for $n_i > k$ we have

$$n_i^{k-1} \frac{(k-1)!}{k^{k-1}} \leq \prod_{t=1}^{k-1} (n_i - t) \leq \frac{F(x, \xi)}{m_F}.$$

Therefore, the inequality in (5.10) holds for $n_i > k$. To conclude that the inequality also holds for $n_i \leq k$, recall that we assumed without loss of generality that $F(y, \xi) > 0$ for all $y \in \xi$, thus $F(x, \xi) > 0$. Hence, it follows from condition (K3) together with symmetry of f that even $F(x, \xi) \geq m_F(k-1)!$. Thus, for $n_i \leq k$, we obtain

$$n_i^{k-1} \frac{(k-1)!}{k^{k-1}} \leq k^{k-1} \frac{(k-1)!}{k^{k-1}} = (k-1)! \leq \frac{F(x, \xi)}{m_F}.$$

This concludes the proof. \square

Proof of Theorem 3.2. The approach that is described in the following to obtain the deviation inequalities presented in Theorem 3.2 is a refinement of the method suggested in [21, 15]. We will use the convex distance for Poisson point processes which was introduced in [19]. Let \mathbf{N}_{fin} be the space of finite point configurations in \mathbb{R}^d , equipped with the σ -algebra \mathcal{N}_{fin} that is obtained by restricting \mathcal{N} to \mathbf{N}_{fin} . Then for any $\xi \in \mathbf{N}_{\text{fin}}$ and $A \in \mathcal{N}_{\text{fin}}$ this distance is given by

$$d_T(\xi, A) = \max_{u \in S(\xi)} \min_{\delta \in A} \sum_{x \in \xi \setminus \delta} u(x),$$

where

$$S(\xi) = \{u : \mathbb{R}^d \rightarrow \mathbb{R}_{\geq 0} \text{ measurable with } \sum_{x \in \xi} u(x)^2 \leq 1\}.$$

To obtain the deviation inequalities for a U-statistic F , we will first relate F in a reasonable way to d_T . Then we will use the inequality

$$(5.11) \quad \mathbb{P}(\eta \in A) \mathbb{P}(d_T(\eta, A) \geq s) \leq \exp\left(-\frac{s^2}{4}\right) \quad \text{for } A \in \mathcal{N}_{\text{fin}}, \ s \geq 0,$$

which was proved in [19]. For the upcoming proof of Theorem 3.2 we also need the following relation, stated in [15]. We carry out the corresponding straightforward computations for the sake of completeness.

Lemma 5.3. *Let F be a U -statistic of order k with kernel $f \geq 0$. Then for any $\xi, \delta \in \mathbf{N}_{\text{fin}}$ we have*

$$F(\xi) \leq k \sum_{x \in \xi \setminus \delta} F(x, \xi) + F(\delta).$$

Proof. We have

$$\begin{aligned} F(\xi) &= \sum_{\mathbf{x} \in \xi_{\neq}^k} \mathbb{1}(\exists x_i \notin \delta) f(\mathbf{x}) + \sum_{\mathbf{x} \in (\xi \cap \delta)_{\neq}^k} f(\mathbf{x}) \\ &\leq \sum_{i=1}^k \sum_{\mathbf{x} \in \xi_{\neq}^k} \mathbb{1}(x_i \notin \delta) f(\mathbf{x}) + \sum_{\mathbf{x} \in \delta_{\neq}^k} f(\mathbf{x}) \\ &= k \sum_{\mathbf{x} \in \xi_{\neq}^k} \mathbb{1}(x_1 \notin \delta) f(\mathbf{x}) + \sum_{\mathbf{x} \in \delta_{\neq}^k} f(\mathbf{x}) \\ &= k \sum_{x \in \xi \setminus \delta} F(x, \xi) + F(\delta) \end{aligned}$$

where the third line holds by symmetry of f . \square

Before we proof Theorem 3.2 in its full generality, we need to establish the corresponding result for finite intensity measure processes. The proof of the next statement is a variation of the method that was suggested in [21, Sections 5.1 and 5.2] and [15, Section 3].

Proposition 5.4. *Assume that the intensity measure of η is finite. Let F be a U -statistic of order k with kernel $f \geq 0$ and let \mathbf{m} be a median of F . Assume that F satisfies (3.1) for some $\alpha \in [0, 2)$ and $c > 0$. Then for all $r \geq 0$ one has*

$$\begin{aligned} \mathbb{P}(F \geq \mathbf{m} + r) &\leq 2 \exp \left(-\frac{r^2}{4k^2 c (r + \mathbf{m})^\alpha} \right), \\ \mathbb{P}(F \leq \mathbf{m} - r) &\leq 2 \exp \left(-\frac{r^2}{4k^2 c \mathbf{m}^\alpha} \right). \end{aligned}$$

Proof. Let $\xi \in \mathbf{N}_{\text{fin}}$ and $A \in \mathcal{N}_{\text{fin}}$. Define the map $u_\xi : \mathbb{R}^d \rightarrow \mathbb{R}_{\geq 0}$ by

$$u_\xi(x) = \frac{F(x, \xi)}{\sqrt{\sum_{y \in \xi} F(y, \xi)^2}} \quad \text{if } x \in \xi \quad \text{and} \quad u_\xi(x) = 0 \quad \text{if } x \notin \xi.$$

Then we have $u_\xi \in S(\xi)$, thus

$$d_T(\xi, A) \geq \min_{\delta \in A} \frac{\sum_{x \in \xi \setminus \delta} F(x, \xi)}{\sqrt{\sum_{y \in \xi} F(y, \xi)^2}}.$$

Moreover, by Lemma 5.3, for any $\delta \in A$ we have

$$F(\xi) \leq k \sum_{x \in \xi \setminus \delta} F(x, \xi) + F(\delta).$$

Thus, since by (3.1) we have $\sum_{y \in \xi} F(y, \xi)^2 \leq cF(\xi)^\alpha$ for \mathbb{P}_η -a.e. $\xi \in \mathbf{N}_{\text{fin}}$, we obtain

$$d_T(\xi, A) \geq \min_{\delta \in A} \frac{F(\xi) - F(\delta)}{k \sqrt{\sum_{y \in \xi} F(y, \xi)^2}} \geq \min_{\delta \in A} \frac{F(\xi) - F(\delta)}{k \sqrt{c} F(\xi)^{\alpha/2}}.$$

Now, to prove the first inequality, let

$$A = \{\delta \in \mathbf{N}_{\text{fin}} : F(\delta) \leq \mathfrak{m}\}.$$

Then, since the map $s \mapsto s/(s + \mathfrak{m})^{\alpha/2}$ is increasing, we have

$$d_T(\xi, A) \geq \frac{F(\xi) - \mathfrak{m}}{k \sqrt{c} F(\xi)^{\alpha/2}} \geq \frac{r}{k \sqrt{c} (r + \mathfrak{m})^{\alpha/2}}$$

for \mathbb{P}_η -a.e. $\xi \in \mathbf{N}_{\text{fin}}$ that satisfies $F(\xi) \geq \mathfrak{m} + r$. This observation together with (5.11) and the fact that $\mathbb{P}(\eta \in A) \geq \frac{1}{2}$ yields

$$\begin{aligned} \mathbb{P}(F(\eta) \geq \mathfrak{m} + r) &\leq \mathbb{P}\left(d_T(\eta, A) \geq \frac{r}{k \sqrt{c} (r + \mathfrak{m})^{\alpha/2}}\right) \\ &\leq 2 \exp\left(-\frac{r^2}{4k^2 c (r + \mathfrak{m})^\alpha}\right). \end{aligned}$$

To prove the second inequality, let $r \geq 0$ and

$$A = \{\delta \in \mathbf{N}_{\text{fin}} : F(\delta) \leq \mathfrak{m} - r\}.$$

If $\mathbb{P}(F(\eta) \leq \mathfrak{m} - r) = \mathbb{P}(\eta \in A) = 0$, the desired inequality holds trivially, so assume that $\mathbb{P}(F(\eta) \leq \mathfrak{m} - r) > 0$ and note that this implies $r \leq \mathfrak{m}$. Then, since the map $s \mapsto (s - (\mathfrak{m} - r))/s^{\alpha/2}$ is increasing, we have

$$d_T(\xi, A) \geq \frac{F(\xi) - (\mathfrak{m} - r)}{k \sqrt{c} F(\xi)^{\alpha/2}} \geq \frac{\mathfrak{m} - (\mathfrak{m} - r)}{k \sqrt{c} \mathfrak{m}^{\alpha/2}} = \frac{r}{k \sqrt{c} \mathfrak{m}^{\alpha/2}}$$

for \mathbb{P}_η -a.e. $\xi \in \mathbf{N}_{\text{fin}}$ that satisfies $F(\xi) \geq \mathfrak{m}$. Thus, it follows from (5.11) that

$$\begin{aligned} \frac{1}{2} &\leq \mathbb{P}(F(\eta) \geq \mathfrak{m}) \leq \mathbb{P}\left(d_T(\eta, A) \geq \frac{r}{k \sqrt{c} \mathfrak{m}^{\alpha/2}}\right) \\ &\leq \frac{1}{\mathbb{P}(F(\eta) \leq \mathfrak{m} - r)} \exp\left(-\frac{r^2}{4k^2 c \mathfrak{m}^\alpha}\right). \end{aligned}$$

□

As a final preparation for the proof of Theorem 3.2, we establish the following lemma. Recall that MX is the smallest median of a random variable X , as defined in (3.2).

Lemma 5.5. *Let X and $X_n, n \in \mathbb{N}$ be random variables such that a.s. $X_{n+1} \geq X_n$ for all $n \in \mathbb{N}$ and $X_n \xrightarrow{\text{a.s.}} X$. Then there exists a non-decreasing sequence $(\mathfrak{m}_n)_{n \in \mathbb{N}}$ where \mathfrak{m}_n is a median of X_n such that $\lim_{n \rightarrow \infty} \mathfrak{m}_n = \text{MX}$.*

Proof. For a random variable Z , let

$$\hat{\text{M}}Z = \sup\{x \in \mathbb{R} : \mathbb{P}(Z \geq x) \geq 1/2\} < \infty.$$

Note that $(\hat{\mathbb{M}}X_n)_{n \in \mathbb{N}}$ is a non-decreasing sequence and that $\hat{\mathbb{M}}X_n \leq \hat{\mathbb{M}}X$ for all $n \in \mathbb{N}$, hence $(\hat{\mathbb{M}}X_n)_{n \in \mathbb{N}}$ is convergent. We claim that

$$(5.12) \quad \lim_{n \rightarrow \infty} \hat{\mathbb{M}}X_n \geq \mathbb{M}X.$$

To see this, let $x \in \mathbb{R}$ be such that $\mathbb{P}(X \leq x) < 1/2$. Since almost sure convergence of the X_n implies convergence in distribution, we have by the Portmanteau theorem (see e.g. [12, Theorem 4.25]) that

$$\limsup_{n \rightarrow \infty} \mathbb{P}(X_n \leq x) \leq \mathbb{P}(X \leq x) < 1/2.$$

Hence, for sufficiently large n , one has $\mathbb{P}(X_n \leq x) < 1/2$. This implies that for sufficiently large n , we have $\mathbb{P}(X_n \geq x) \geq 1/2$ and thus $\hat{\mathbb{M}}X_n \geq x$. From these considerations it follows that

$$\begin{aligned} \lim_{n \rightarrow \infty} \hat{\mathbb{M}}X_n &\geq \sup\{x \in \mathbb{R} : \mathbb{P}(X \leq x) < \tfrac{1}{2}\} \\ &= \inf\{x \in \mathbb{R} : \mathbb{P}(X \leq x) \geq \tfrac{1}{2}\} = \mathbb{M}X. \end{aligned}$$

Hence (5.12) is established. Now, for any $n \in \mathbb{N}$, either $\mathbb{M}X_n = \hat{\mathbb{M}}X_n$ is the unique median of X_n or all elements in the interval $[\mathbb{M}X_n, \hat{\mathbb{M}}X_n]$ are medians of X_n , where $\mathbb{M}X_n$ and $\hat{\mathbb{M}}X_n$ are non-decreasing in n . Taking (5.12) into account as well as the fact that $\mathbb{M}X_n \leq \mathbb{M}X$ for all $n \in \mathbb{N}$, the result follows. \square

Proof of Theorem 3.2. For any $n \in \mathbb{N}$ let $\eta_n = \eta \cap B(0, n)$. Then η_n is a Poisson point process with finite intensity measure that is given by $\mu_n(A) = \mu(A \cap B(0, n))$ for any Borel set $A \subseteq \mathbb{R}^d$. We define for any $n \in \mathbb{N}$ the random variable F_n in terms of the functional $F : \mathbf{N} \rightarrow \mathbb{R} \cup \{\infty\}$ by $F_n = F(\eta_n)$. One easily observes that, since F is a U-statistic with non-negative kernel, we have almost surely $F_{n+1} \geq F_n$ for all $n \in \mathbb{N}$ and also $F_n \xrightarrow{a.s.} F$. According to Lemma 5.5 we can choose a non-decreasing sequence $\mathbf{m}_n, n \in \mathbb{N}$ such that \mathbf{m}_n is a median of F_n satisfying $\lim_{n \rightarrow \infty} \mathbf{m}_n = \mathbb{M}F$. Let $r \geq 0$. By virtue of Proposition 5.4, for any $n \in \mathbb{N}$ we have

$$\mathbb{P}(F_n - \mathbf{m}_n \geq r) \leq 2 \exp\left(-\frac{r^2}{4k^2c(r + \mathbf{m}_n)^\alpha}\right) \leq 2 \exp\left(-\frac{r^2}{4k^2c(r + \mathbb{M}F)^\alpha}\right).$$

The sequence of random variables $F_n - \mathbf{m}_n$ converges almost surely (and thus in distribution) to $F - \mathbb{M}F$. Therefore, by the Portmanteau theorem (see e.g. [12, Theorem 4.25]) we have

$$\mathbb{P}(F - \mathbb{M}F > r) \leq \liminf_{n \rightarrow \infty} \mathbb{P}(F_n - \mathbf{m}_n > r) \leq 2 \exp\left(-\frac{r^2}{4k^2c(r + \mathbb{M}F)^\alpha}\right).$$

The inequality for the lower tail follows analogously. \square

Proof of Proposition 3.5. The proof presented below generalizes ideas from [1, Section 6.1]. Recall that F is a U-statistic of order k satisfying (K1) to (K3) and that I and $a > 0$ are chosen such that

$$(3.4) \quad \mathbb{P}(F \geq M + r) \leq \exp(-I(r)),$$

where M is either the mean or a median of F , and

$$(3.5) \quad \liminf_{r \rightarrow \infty} I(r)/r^a > 0.$$

Proof of Proposition 3.5. Choose some $x \in \mathbb{R}^d$ such that $q = \mu(B(x, \rho_F/2)) > 0$. Then $Z := \eta(B(x, \rho_F/2))$ is a Poisson random variable with mean q . Since the diameter of $B(x, \rho_F/2)$ equals ρ_F , the assumptions (K1) and (K3) yield that for any $\mathbf{x} \in \eta_{\neq}^k \cap B(x, \rho_F/2)$ one has $f(\mathbf{x}) \geq m_F > 0$. It follows that

$$F \geq m_F k! \binom{Z}{k}.$$

Hence, there exists a constant $A > 0$ such that for sufficiently large r ,

$$AZ^k \geq r \text{ implies } F \geq r.$$

Thus, for sufficiently large r ,

$$(5.13) \quad \mathbb{P}(F \geq M + r) \geq \mathbb{P}(AZ^k \geq M + r) = \mathbb{P}(Z \geq \tau(r)),$$

where $\tau(r) = A^{-1/k}(r + M)^{1/k}$. The tail asymptotic of a Poisson random variable is well known and one has $\mathbb{P}(Z \geq s) \sim \exp(-s \log(s/q) - q)$ as $s \rightarrow \infty$, see e.g. [9]. Here the symbol \sim denotes the asymptotic equivalence relation. Combining this with (5.13), we obtain

$$\liminf_{r \rightarrow \infty} \frac{\mathbb{P}(F \geq M + r)}{\exp(-\tau(r) \log(\tau(r)/q) - q)} \geq 1.$$

Hence, taking (3.4) into account, there exists a constant $B > 0$ such that for sufficiently large r ,

$$\tau(r) \log(\tau(r)/q) + q \geq I(r) - B.$$

By virtue of (3.5), dividing this inequality by r^a and taking the limit yields

$$(5.14) \quad \liminf_{r \rightarrow \infty} \frac{\tau(r) \log(\tau(r)/q)}{r^a} > 0.$$

Moreover, writing $C = A^{-1/k} > 0$, we have

$$\frac{\tau(r) \log(\tau(r)/q)}{r^a} \sim C r^{1/k-a} \log(C(r + M)^{1/k}/q).$$

The result follows since $a > 1/k$ would imply that the RHS in the above display converges to 0, contradicting (5.14). \square

Proof of Proposition 4.1 and Theorem 4.2. The upcoming reasoning is partially inspired by the proof of [18, Proposition 3.1]. First of all, we need to introduce the quantities $\|f_n\|_n^2$ that are crucial ingredients for the stochastic analysis of Poisson processes using Malliavin Calculus, see e.g. [16] for details. According to [20, Lemma 3.5], for a square-integrable U-statistic with kernel f these quantities can be explicitly written as

$$(5.15) \quad \|f_n\|_n^2 = \binom{k}{n}^2 \int_{(\mathbb{R}^d)^n} \left(\int_{(\mathbb{R}^d)^{k-n}} f(\mathbf{y}_n, \mathbf{x}_{k-n}) d\mu^{k-n}(\mathbf{x}_{k-n}) \right)^2 d\mu^n(\mathbf{y}_n),$$

where we use the abbreviations $\mathbf{y}_n = (y_1, \dots, y_n)$ and $\mathbf{x}_{k-n} = (x_{n+1}, \dots, x_k)$. Also according to [20, Lemma 3.5], one has that the variance of a square-integrable U-statistic F is now given by

$$(5.16) \quad \mathbb{V}F = \sum_{n=1}^k n! \|f_n\|_n^2.$$

This formula will be crucial for analysing the variance asymptotics of subgraph counts. For the remainder of the section, let the assumptions and notations of Section 4.2 prevail. In particular, recall that m denotes the Lebesgue density of the measure μ and note that we will use the notation

$$m^{\otimes l}(\mathbf{z}) = m^{\otimes l}(z_1, \dots, z_l) = \prod_{i=1}^l m(z_i)$$

for any $l \in \mathbb{N}$ and $\mathbf{z} = (z_1, \dots, z_l) \in (\mathbb{R}^d)^l$. The following two lemmas are used in the proofs of Proposition 4.1 and Theorem 4.2.

Lemma 5.6. *Assume that the subgraph count N_t is square-integrable. Then the corresponding quantities $\|f_n\|_n^2$ given by formula (5.15) can be written as*

$$(5.17) \quad c_n(t) \int_{\mathbb{R}^d \times B(0, \Theta)^{n-1}} I_t^y(\mathbf{y}_n) \left(\int_{B(0, \Theta)^{k-n}} I_t^x(\mathbf{x}_{k-n}) J(\mathbf{y}_n, \mathbf{x}_{k-n}) d\mathbf{x}_{k-n} \right)^2 d\mathbf{y}_n,$$

where

$$c_n(t) = \frac{t^{2k-n} \rho_t^{d(2k-n-1)}}{(k!)^2} \binom{k}{n}^2,$$

$$\Theta = \text{diam}(H)\theta,$$

$$I_t^y(\mathbf{y}_n) = m^{\otimes n}(y_1, \rho_t y_2 + y_1, \dots, \rho_t y_n + y_1),$$

$$I_t^x(\mathbf{x}_{k-n}) = m^{\otimes k-n}(\rho_t x_{n+1} + y_1, \dots, \rho_t x_k + y_1),$$

$$J(\mathbf{y}_n, \mathbf{x}_{k-n}) = |\{\text{subgraphs } H' \text{ of } G_S(0, y_2, \dots, y_n, x_{n+1}, \dots, x_k) : H' \cong H\}|.$$

Proof. Recall that the kernel of the subgraph count N_t is given by

$$f_t(\mathbf{z}) = \frac{|\{\text{subgraphs } H' \text{ of } G_{\rho_t S}(\mathbf{z}) : H' \cong H\}|}{k!}, \quad \mathbf{z} \in (\mathbb{R}^d)^k.$$

Since N_t is assumed to be square-integrable, formula (5.15) holds. The integral on the RHS of (5.15) can be written as

$$t^{2k-n} \int_{(\mathbb{R}^d)^n} m^{\otimes n}(\mathbf{y}_n) \left(\int_{(\mathbb{R}^d)^{k-n}} m^{\otimes k-n}(\mathbf{x}_{k-n}) f_t(\mathbf{y}_n, \mathbf{x}_{k-n}) d\mathbf{x}_{k-n} \right)^2 d\mathbf{y}_n.$$

The change of variables

$$\mathbf{x}_{k-n} \mapsto (\rho_t x_{n+1} + y_1, \dots, \rho_t x_k + y_1)$$

now gives that the above expression equals

$$\frac{t^{2k-n} \rho_t^{2d(k-n)}}{(k!)^2} \int_{(\mathbb{R}^d)^n} m^{\otimes n}(\mathbf{y}_n) \left(\int_{(\mathbb{R}^d)^{k-n}} I_t^x(\mathbf{x}_{k-n}) J_t^x(\mathbf{y}_n, \mathbf{x}_{k-n}) d\mathbf{x}_{k-n} \right)^2 d\mathbf{y}_n,$$

where

$$J_t^x(\mathbf{y}_n, \mathbf{x}_{k-n}) = |\{\text{subgraphs } H' \text{ of } G_{\rho_t S}(\mathbf{y}_n, \rho_t x_{n+1} + y_1, \dots, \rho_t x_k + y_1) : H' \cong H\}|.$$

We perform a further change of variables for the outer integral, namely

$$\mathbf{y}_n \mapsto (y_1, \rho_t y_2 + y_1, \dots, \rho_t y_n + y_1).$$

This yields that the quantity in question equals

$$\frac{t^{2k-n} \rho_t^{d(2k-n-1)}}{(k!)^2} \int_{(\mathbb{R}^d)^n} I_t^y(\mathbf{y}_n) \left(\int_{(\mathbb{R}^d)^{k-n}} I_t^x(\mathbf{x}_{k-n}) J(\mathbf{y}_n, \mathbf{x}_{k-n}) d\mathbf{x}_{k-n} \right)^2 d\mathbf{y}_n,$$

where one should notice that

$$\begin{aligned} & J_t^x(y_1, \rho_t y_2 + y_1, \dots, \rho_t y_n + y_1, \mathbf{x}_{k-n}) \\ &= |\{\text{subgraphs } H' \text{ of } G_S(0, y_2, \dots, y_n, x_{n+1}, \dots, x_k) : H' \cong H\}| \\ &= J(\mathbf{y}_n, \mathbf{x}_{k-n}). \end{aligned}$$

Also, we see that $J(\mathbf{y}_n, \mathbf{x}_{k-n}) = 0$ whenever $\|x_i\| > \text{diam}(H)\theta = \Theta$ for some $n+1 \leq i \leq k$. Hence, the inner integral is actually an integral over $B(0, \Theta)^{k-n}$. Similarly, the outer integral is an integral over $\mathbb{R}^d \times B(0, \Theta)^{n-1}$. \square

In the next statement we use the abbreviation $\mathbf{x} = (x_1, \dots, x_k)$. Note that here we do *not* assume that the considered subgraph count N_t is necessarily square-integrable.

Lemma 5.7. *The expectation of N_t can be written as*

$$(5.18) \quad \mathbb{E}N_t = \frac{t^k \rho_t^{d(k-1)}}{k!} \int_{\mathbb{R}^d \times B(0, \Theta)^{k-1}} I_t(\mathbf{x}) J(\mathbf{x}) d\mathbf{x},$$

where

$$(5.19) \quad \begin{aligned} I_t(\mathbf{x}) &= m^{\otimes k}(x_1, \rho_t x_2 + x_1, \dots, \rho_t x_k + x_1), \\ J(\mathbf{x}) &= |\{\text{subgraphs } H' \text{ of } G_S(0, x_2, \dots, x_k) : H' \cong H\}|. \end{aligned}$$

In (5.18), the LHS is finite if and only if the RHS is finite.

Proof. Using the Slivniak-Mecke formula (see e.g. [22, Corollary 3.2.3]) we can write

$$\mathbb{E}N_t = \frac{t^k}{k!} \int_{(\mathbb{R}^d)^k} m^{\otimes k}(\mathbf{x}) |\{\text{subgraphs } H' \text{ of } G_{\rho_t S}(\mathbf{x}) : H' \cong H\}| d\mathbf{x}.$$

Now, the change of variables

$$\mathbf{x} = (x_1, \dots, x_k) \mapsto (x_1, \rho_t x_2 + x_1, \dots, \rho_t x_k + x_1)$$

followed by a reasoning very similar to the one in the proof of Lemma 5.6 yields the result. \square

Proof of Proposition 4.1. Let the notation of Lemma 5.7 prevail. The assumption (4.6) implies that for any $t \in \mathbb{N}$ and $\mathbf{x} \in \mathbb{R}^d \times B(0, \Theta)^{k-1}$,

$$c^{-k}m(x_1)^k \leq I_t(\mathbf{x}) \leq c^k m(x_1)^k.$$

Using Lemma 5.7, one obtains

$$c^{-k}A \int_{\mathbb{R}^d} m(x)^k dx \leq \frac{k!}{t^k \rho_t^{d(k-1)}} \mathbb{E}N_t \leq c^k A \int_{\mathbb{R}^d} m(x)^k dx,$$

where

$$A = \int_{B(0, \Theta)^{k-1}} |\{\text{subgraphs } H' \text{ of } G_S(0, x_2, \dots, x_k) : H' \cong H\}| d\mathbf{x} > 0.$$

Hence, it follows that $\int m(x)^k dx < \infty$ is equivalent to integrability of N_t . According to Theorem 1.1, the subgraph count N_t is integrable if and only if it is almost surely finite. The result follows. \square

Proof of Theorem 4.2. [Proof of (i)] Let the notation of Lemma 5.7 prevail. Since by assumption the density m is continuous, we have for I_t given by (5.19) and for any $\mathbf{x} \in (\mathbb{R}^d)^k$,

$$\lim_{t \rightarrow \infty} I_t(\mathbf{x}) = m(x_1)^k.$$

Now, the condition (4.6) guarantees that for any $\mathbf{x} \in \mathbb{R}^d \times B(0, \Theta)^{k-1}$,

$$I_t(\mathbf{x}) \leq c^k m(x_1)^k.$$

Moreover, by virtue of assumption (4.7) and boundedness of J ,

$$K := \int_{\mathbb{R}^d \times B(0, \Theta)^{k-1}} m(x_1)^k J(\mathbf{x}) d\mathbf{x} < \infty.$$

Hence, Lemma 5.7 together with the dominated convergence theorem yield that the sequence $\mathbb{E}N_t / (t^k \rho_t^{d(k-1)})$ converges to $K/k! > 0$.

[Proof of (ii)] Since $\mathbb{M}N_t$ is a median of N_t , in the case $\mathbb{M}N_t > \mathbb{E}N_t$ one has

$$\begin{aligned} \frac{1}{2} &\leq \mathbb{P}(N_t \geq \mathbb{M}N_t) = \mathbb{P}(N_t - \mathbb{E}N_t \geq \mathbb{M}N_t - \mathbb{E}N_t) \\ &\leq \mathbb{P}(|N_t - \mathbb{E}N_t| \geq |\mathbb{M}N_t - \mathbb{E}N_t|) \leq \frac{\mathbb{V}N_t}{(\mathbb{M}N_t - \mathbb{E}N_t)^2}, \end{aligned}$$

where for the last inequality we used Chebyshev. Similarly, if $\mathbb{M}N_t < \mathbb{E}N_t$,

$$\frac{1}{2} \leq \mathbb{P}(N_t \leq \mathbb{M}N_t) \leq \mathbb{P}(|N_t - \mathbb{E}N_t| \geq |\mathbb{M}N_t - \mathbb{E}N_t|) \leq \frac{\mathbb{V}N_t}{(\mathbb{M}N_t - \mathbb{E}N_t)^2}.$$

We see that in any case one has

$$\left| \frac{\mathbb{M}N_t}{\mathbb{E}N_t} - 1 \right| \leq \frac{\sqrt{2\mathbb{V}N_t}}{\mathbb{E}N_t}.$$

Now, the statements (i) and (iii) of the theorem yield

$$(5.20) \quad \frac{\sqrt{2\mathbb{V}N_t}}{\mathbb{E}N_t} \sim \sqrt{\frac{2}{a^2} \sum_{n=1}^k \frac{(t\rho_t^d)^{k-n}}{t^k \rho_t^{d(k-1)}} A^{(n)}} = \sqrt{\frac{2}{a^2} \sum_{n=1}^k \rho_t^{d(1-n/k)} (t^k \rho_t^{d(k-1)})^{-n/k} A^{(n)}}.$$

The assumption $\lim_{t \rightarrow \infty} \mathbb{E}N_t = \infty$ together with statement (i) of the theorem and $\lim_{t \rightarrow \infty} \rho_t = 0$ implies that this tends to 0 as $t \rightarrow \infty$.

[Proof of (iii)] First note that by Proposition 4.1 together with Theorem 1.1 the assumption (4.7) guarantees that all moments of the subgraph counts N_t exist. In particular, the N_t are square-integrable, thus formula (5.16) is in order. So, to analyse the asymptotic behaviour of $\mathbb{V}N_t$, we need to analyse the asymptotics of the quantities $\|f_n\|_n^2$. This will be done by proving convergence of the integrals that appear in (5.17),

$$\int_{\mathbb{R}^d \times B(0, \Theta)^{n-1}} I_t^y(\mathbf{y}_n) \left(\int_{B(0, \Theta)^{k-n}} I_t^x(\mathbf{x}_{k-n}) J(\mathbf{y}_n, \mathbf{x}_{k-n}) d\mathbf{x}_{k-n} \right)^2 d\mathbf{y}_n.$$

We first remark that by continuity of the density m , for any $\mathbf{x}_{k-n} \in (\mathbb{R}^d)^{k-n}$,

$$\lim_{t \rightarrow \infty} I_t^x(\mathbf{x}_{k-n}) = m(y_1)^{k-n}.$$

Since J and m are bounded, the dominated convergence theorem applies and gives

$$\begin{aligned} & \lim_{t \rightarrow \infty} \int_{B(0, \Theta)^{k-n}} I_t^x(\mathbf{x}_{k-n}) J(\mathbf{y}_n, \mathbf{x}_{k-n}) d\mathbf{x}_{k-n} \\ &= \int_{B(0, \Theta)^{k-n}} m(y_1)^{k-n} J(\mathbf{y}_n, \mathbf{x}_{k-n}) d\mathbf{x}_{k-n}. \end{aligned}$$

Now, it follows from the assumption (4.6) that for any $\mathbf{x}_{k-n} \in B(0, \Theta)^{k-n}$ and $\mathbf{y}_n \in \mathbb{R}^d \times B(0, \Theta)^{n-1}$,

$$I_t^x(\mathbf{x}_{k-n}) \leq c^{k-n} m(y_1)^{k-n} \quad \text{and} \quad I_t^y(\mathbf{y}_n) \leq c^n m(y_1)^n.$$

Moreover, we have that

$$\begin{aligned} & \int_{\mathbb{R}^d \times B(0, \Theta)^{n-1}} m(y_1)^n \left(\int_{B(0, \Theta)^{k-n}} m(y_1)^{k-n} J(\mathbf{y}_n, \mathbf{x}_{k-n}) d\mathbf{x}_{k-n} \right)^2 d\mathbf{y}_n \\ & \leq (\Theta^d \kappa_d)^{2k-n-1} \|J\|_\infty^2 \int_{\mathbb{R}^d} m(x)^{2k-n} dx < \infty, \end{aligned}$$

where κ_d denotes the Lebesgue measure of the unit ball in \mathbb{R}^d . Here the finiteness of the latter integral follows from assumption (4.7) since m is bounded and $2k-n \geq k$. Furthermore, by continuity of m , one has for any $\mathbf{y}_n \in \mathbb{R}^d \times B(0, \Theta)^{n-1}$,

$$\lim_{t \rightarrow \infty} I_t^y(\mathbf{y}_n) = m(y_1)^n.$$

We conclude that the dominated convergence theorem applies and gives

$$\begin{aligned}
& \lim_{t \rightarrow \infty} \int_{\mathbb{R}^d \times B(0, \Theta)^{n-1}} I_t^y(\mathbf{y}_n) \left(\int_{B(0, \Theta)^{k-n}} I_t^x(\mathbf{x}_{k-n}) J(\mathbf{y}_n, \mathbf{x}_{k-n}) d\mathbf{x}_{k-n} \right)^2 d\mathbf{y}_n \\
&= \int_{\mathbb{R}^d} m(x)^{2k-n} dx \int_{B(0, \Theta)^{n-1}} \left(\int_{B(0, \Theta)^{k-n}} J^0(\mathbf{y}_{n-1}, \mathbf{x}_{k-n}) d\mathbf{x}_{k-n} \right)^2 d\mathbf{y}_{n-1} \\
&=: K^{(n)},
\end{aligned}$$

where

$$J^0(\mathbf{y}_{n-1}, \mathbf{x}_{k-n}) = |\{\text{subgraphs } H' \text{ of } G_S(0, \mathbf{y}_{n-1}, \mathbf{x}_{k-n}) : H' \cong H\}|.$$

Hence, by virtue of (5.16) and (5.17) the variance can be written as

$$\begin{aligned}
\mathbb{V}N_t &= \sum_{n=1}^k t^{2k-n} \rho_t^{d(2k-n-1)} \frac{n!}{(k!)^2} \binom{k}{n}^2 K_t^{(n)}, \\
&= t^k \rho_t^{d(k-1)} \sum_{n=1}^k (t \rho_t^d)^{k-n} \frac{n!}{(k!)^2} \binom{k}{n}^2 K_t^{(n)}.
\end{aligned}$$

where $K_t^{(n)}$ is such that $\lim_{t \rightarrow \infty} K_t^{(n)} = K^{(n)} \in (0, \infty)$. The result follows. \square

Proof of Theorem 4.3. In the upcoming proof of the strong law for subgraph counts we use the following consequence of the Borel-Cantelli lemma (see e.g. [12, Theorem 3.18]), well known from basic probability theory.

Lemma 5.8. *Let $(X_n)_{n \in \mathbb{N}}$ be a sequence of real random variables and let $(a_n)_{n \in \mathbb{N}}$ be a sequence of real numbers converging to some $a \in \mathbb{R}$. Assume that for any $\varepsilon > 0$,*

$$\sum_{n=1}^{\infty} \mathbb{P}(|X_n - a_n| \geq \varepsilon) < \infty.$$

Then

$$X_n \xrightarrow{a.s.} a \quad \text{as } n \rightarrow \infty.$$

Proof of Theorem 4.3. It follows from Theorem 1.1 that for any real number $\varepsilon > 0$,

$$\mathbb{P} \left(\left| \frac{N_t}{t^k \rho_t^{d(k-1)}} - \frac{\mathbb{E}N_t}{t^k \rho_t^{d(k-1)}} \right| \geq \varepsilon \right) \leq p_u(t, \varepsilon) + p_l(t, \varepsilon),$$

where

$$p_u(t, \varepsilon) := \exp \left(-t \rho_t^{d(k-1)/k} \frac{\left(\left(\frac{\mathbb{E} N_t}{t^k \rho_t^{d(k-1)}} + \varepsilon \right)^{\frac{1}{2k}} - \left(\frac{\mathbb{E} N_t}{t^k \rho_t^{d(k-1)}} \right)^{\frac{1}{2k}} \right)^2}{2k^2 c_d} \right),$$

$$p_l(t, \varepsilon) := \exp \left(-\frac{t^{2k} \rho_t^{2d(k-1)} \varepsilon^2}{2k \mathbb{V} N_t} \right).$$

Now, by Theorem 4.2 (i) the sequence $\mathbb{E} N_t / (t^k \rho_t^{d(k-1)})$ converges to a positive real number. This implies that we can choose a constant $C > 0$ such that for sufficiently large t ,

$$p_u(t, \varepsilon) \leq \exp \left(-t \rho_t^{d(k-1)/k} C \right).$$

Moreover, assumption (4.8) guarantees the existence of a constant $C' > 0$ such that for sufficiently large t ,

$$\exp \left(-t \rho_t^{d(k-1)/k} C \right) \leq \exp \left(-t^{\gamma/k} C' \right).$$

We conclude that $\sum_{t \in \mathbb{N}} p_u(t, \varepsilon) < \infty$. Furthermore, similarly as in (5.20), it follows from Theorem 4.2 (iii) that

$$\frac{\mathbb{V} N_t}{t^{2k} \rho_t^{2d(k-1)}} \sim \sum_{n=1}^k \rho_t^{d(1-n/k)} (t^k \rho_t^{d(k-1)})^{-n/k} A^{(n)}.$$

Invoking assumption (4.8) yields existence of a constant $C'' > 0$ such that for sufficiently large t ,

$$\sum_{n=1}^k \rho_t^{d(1-n/k)} (t^k \rho_t^{d(k-1)})^{-n/k} A^{(n)} \leq \sum_{n=1}^k \rho_t^{d(1-n/k)} (C'' t^{\gamma})^{-n/k} A^{(n)}.$$

Now, the sequence ρ_t tends to 0 and thus it is bounded above. Moreover, for any $t \in \mathbb{N}$ one has $t^{-k\gamma/k} \leq t^{-(k-1)\gamma/k} \leq \dots \leq t^{-\gamma/k}$. From this together with the last two displays we derive existence of a constant $C''' > 0$ such that for large t ,

$$\frac{\mathbb{V} N_t}{t^{2k} \rho_t^{2d(k-1)}} \leq C''' t^{-\gamma/k}.$$

Hence, for sufficiently large t ,

$$p_l(t, \varepsilon) \leq \exp \left(-\frac{t^{\gamma/k} \varepsilon^2}{2k C'''} \right).$$

It follows that $\sum_{t \in \mathbb{N}} p_l(t, \varepsilon) < \infty$. Applying Lemma 5.8 yields the result. \square

REFERENCES

- [1] S. Bachmann and G. Peccati. *Concentration Bounds for Geometric Poisson Functionals: Logarithmic Sobolev Inequalities Revisited*. Version 1. Apr. 13, 2015. arXiv:1504.03138 [math.PR].
- [2] O. Bobrowski and M. Kahle. *Topology of random geometric complexes: a survey*. Version 1. Sept. 16, 2014. arXiv:1409.4734 [math.AT].
- [3] G. Carlsson. “Topology and data”. In: *Bull. Amer. Math. Soc. (N.S.)* 46.2 (2009), pp. 255–308.
- [4] G. Chen, C. Yao, and T. Guo. “The asymptotic size of the largest component in random geometric graphs with some applications”. In: *Adv. in Appl. Probab.* 46.2 (2014), pp. 307–324.
- [5] L. Decreusefond and E. Ferraz. “On the one dimensional Poisson random geometric graph”. In: *J. Probab. Stat.* (2011), Art. ID 350382, 21.
- [6] L. Decreusefond, E. Ferraz, H. Randriambololona, and A. Vergne. “Simplicial homology of random configurations”. In: *Adv. in Appl. Probab.* 46.2 (2014), pp. 325–347.
- [7] P. Eichelsbacher, M. Raič, and T. Schreiber. “Moderate deviations for stabilizing functionals in geometric probability”. In: *Ann. Inst. Henri Poincaré Probab. Stat.* 51.1 (2015), pp. 89–128.
- [8] E. N. Gilbert. “Random plane networks”. In: *J. Soc. Indust. Appl. Math.* 9 (1961), pp. 533–543.
- [9] P. W. Glynn. “Upper bounds on Poisson tail probabilities”. In: *Operations Research Letters* 6.1 (1987), pp. 9–14.
- [10] M. Kahle. “Random geometric complexes”. In: *Discrete Comput. Geom.* 45.3 (2011), pp. 553–573.
- [11] M. Kahle and E. Meckes. “Limit theorems for Betti numbers of random simplicial complexes”. In: *Homology Homotopy Appl.* 15.1 (2013), pp. 343–374.
- [12] O. Kallenberg. *Foundations of modern probability*. Second. Probability and its Applications (New York). Springer-Verlag, New York, 2002, pp. xx+638.
- [13] R. Lachièze-Rey and G. Peccati. “Fine Gaussian fluctuations on the Poisson space, I: contractions, cumulants and geometric random graphs”. In: *Electron. J. Probab.* 18 (2013), no. 32, 32.
- [14] R. Lachièze-Rey and G. Peccati. “Fine Gaussian fluctuations on the Poisson space II: rescaled kernels, marked processes and geometric U -statistics”. In: *Stochastic Process. Appl.* 123.12 (2013), pp. 4186–4218.
- [15] R. Lachièze-Rey and M. Reitzner. *U -statistics in stochastic geometry*. Version 2. Mar. 26, 2015. arXiv:1503.00110v2 [math.PR].
- [16] G. Last. *Stochastic analysis for Poisson processes*. Version 1. May 17, 2014. arXiv:1405.4416 [math.PR].
- [17] S. Muthukrishnan and G. Pandurangan. “Thresholding random geometric graph properties motivated by ad hoc sensor networks”. In: *J. Comput. System Sci.* 76.7 (2010), pp. 686–696.
- [18] M. D. Penrose. *Random geometric graphs*. Vol. 5. Oxford Studies in Probability. Oxford: Oxford University Press, 2003, pp. xiv+330.

- [19] M. Reitzner. “Poisson point processes: large deviation inequalities for the convex distance”. In: *Electron. Commun. Probab.* 18 (2013), no. 96, 7.
- [20] M. Reitzner and M. Schulte. “Central limit theorems for U -statistics of Poisson point processes”. In: *Ann. Probab.* 41.6 (2013), pp. 3879–3909.
- [21] M. Reitzner, M. Schulte, and Ch. Thäle. *Limit theory for the Gilbert graph*. Version 1. Oct. 17, 2013. arXiv:1312.4861 [math.PR].
- [22] R. Schneider and W. Weil. *Stochastic and integral geometry*. Probability and its Applications (New York). Springer-Verlag, Berlin, 2008, pp. xii+693.
- [23] V. de Silva and R. Ghrist. “Coverage in sensor networks via persistent homology”. In: *Algebr. Geom. Topol.* 7 (2007), pp. 339–358.
- [24] D. Yogeshwaran, E. Subag, and R. J. Adler. *Random geometric complexes in the thermodynamic regime*. Version 1. Mar. 5, 2014. arXiv:1403.1164 [math.PR].

SASCHA BACHMANN
INSTITUT FÜR MATHEMATIK
UNIVERSITÄT OSNABRÜCK
49069 OSNABRÜCK, GERMANY
EMAIL: SASCHA.BACHMANN@UNI-OSNABRUECK.DE

MATTHIAS REITZNER
INSTITUT FÜR MATHEMATIK
UNIVERSITÄT OSNABRÜCK
49069 OSNABRÜCK, GERMANY
EMAIL: MATTHIAS.REITZNER@UNI-OSNABRUECK.DE